Problem Statement: Bank of Portugal Telephone Campaign Predicting Model

Jess Zhang

April 9, 2018

Project Background

Bank of Portugal uses telemarketing campaigns to generate term deposits. The conversion rate affects the profit margin of the campaign directly. Bank of Portugal would like to work with data scientists from CSX415 to develop a machine learning model to predict the outcome of future campaigns. With the new model, Bank of Portugal can better plan their outgoing calls and target its client base more effectively.

The campaign department of Bank of Portugal has collected its past campaign data for this purpose. The data includes both clients profile information such as age, job, education, etc. The data also contains calling details such as number of calls, final results, call duration, etc. In addition, the data recorded economic data such as unemployment rate.

Stakeholders

The stakeholders who will be involved in this project are:

- John Smith – Campaign Manager from Bank of Portugal's Telemarketing department. He is the point of contact for Bank of Portugal. John and his team will be using the new model to analyze their client base and develop campaign strategies.
Email: John.Smith@bancdeportugal.com

- Ashley Harvey – Data Analyst from Bank of Portugal. Her job is to collect and record campaign data. Once the new model is deployed, Ashley will be the primary user of this model and to report results to John. Ashley is proficient with R programming and is able to run R scripts and packages. Email: Ashley.Harvey@bancdeportugal.com

- Jessica Zhang – Data Scientist from CSX415. She is the assigned to this project to develop an effective machine learning model to predict campaign outcome therefore help Bank of Portugal to improve their campaign effectiveness. Email: Jesszh@Gmail.com

- Nick Nolasco – Data Engineer from CSX415. He is assigned to this project to analyze and clean the dataset provided by Bank of Portugal so the data scientist can analyze and train the data efficiently. Email: Nick.N@gmail.com

Project Scope

Many factors, such as age, job, and loan amount can affect people's decisions on a term deposit. The goal of this project is to provide Bank of Portugal with useful prediction for future campaigns.

The current campaign success rate from this dataset is at 23.6%, which means more than 75 percent of calls yield no deposit. In order to better utilize staffs' time, the model has to be able to predict better than 77 percent to be useful.

Developing of the marketing strategy is outside of the scope of this project.

A few algorithms will be run and the best suited model will be built to analyze the given dataset. The CSX425 team will share and disucss the analysis with the campaign manager of Bank of Portugal. The campaign manager and his team can continue to use the model to analyze data in the future.

The delivered model will work with new data with the same format in the datafile. However, the model may not work if the client adds or removes columns from the datafile without counseling the CSX415 project team first.

Success Criteria

The average probability of each campaign call is 23%, a blanket "no" prediction has 77 percent chance to be right. Bank of Portugal and the CSX415 project team agree that a model with over 90 percent accuracy will be efficient.

Benefits of the Project

With a model which can predict at least 90 percent of the outcome, the campaign manager can focus most his team's effort on the clients who are most likely to make a term deposit during busy season and call the rest during low season.

The campaign team can also develop different call tactics towards different profile group or during a specific time period to improve the overall campaign effectiveness.

With future data gathering, Bank of Portugal can continue to train the model and improve the prediction accuracy. The campaign manager can therefore further improve his strategy.

Estimated Risks

It is reasonable to believe that there are patterns can be generated from the given dataset. It is unlikely however possible that no meaningful pattern can be identified from given data due to insufficient data.

Delivery Timeline

The CSX415 project team will be working on this project for the next 2 months. One data scientist and one data engineer will be working on the model and analysis full-time. Below is the schedule of the deliverables:

Milestone one: 4/30/2018

- The CSX415 team will deliver the preliminary result to the client on 4/30/2018.

- The CSX415 team will discuss the final delivery date with Bank of Portugal.

- The project scope and final delivery date can be adjusted based on the result of deliverable one, upon the mutual agreement of both CSX415 team and Bank of Portugal.

Final Delivery: 5/31/2018

- The estimated final delivery date is May 31st, 2018. The CSX415 team will give a final demo to the campaign team from Bank of Portugal on this date.

- The CSX415 team will transfer the entire code base and analysis result to Bank of Portugal.

- The CSX415 team will train one representative data analyst from Bank of Portugal on the usage and the basic maintenance of the model.

- The CSX415 team will provide detailed instructions on how to train and maintain the model.

Client Responsibilities After Delivery

Bank of Portugal should run and maintain the model after the initial deployment. The CSX415 project team agrees to review the model upon request by Bank of Portugal.

Bank of Portugal is responsible for updating the data and retrain the model.

Do not change the format of the data or the model may not run again!

Since the users of this model may not have sufficient knowledge about machine learning, it is possible for someone to accidentally remove or damage the model or one of the supporting files. Weekly backup of the data as well as all files from this deliverable is highly recommended.

The data model's performance will improve with more data gathering and training. However, after a period of time, the new data may be different from the original data enough that the difference will affect the model performance negatively. CSX41 recommends Bank of Portugal to bring the model back for performance tuning once the model performance decreases more than 10 percent, or annually, whichever comes first.

Data Source

Bank Marketing Data is taken from http://archive.ics.uci.edu/ml/datasets/Bank+Marketing and placed under /data folder.